DOCUMENT RESUME

ED 348 383 TM 018 762

AUTHOR Fan, Xitao
TITLE Canonical Correlation Analysis as a General Analytical Model.
PUB DATE Apr 92
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Chi Square; Classification; Computer Oriented Programs; *Correlation; *Discriminant Analysis; Equations (Mathematics); *Mathematical Models; *Multivariate Analysis
IDENTIFIERS *Contingency Tables

ABSTRACT

This paper focuses on three aspects related to the conceptualization and application of canonical correlation analysis as a dominant statistical model: (1) partial canonical correlation analysis and its application in statistical testing; (2) the relation between canonical correlation analysis and discriminant analysis; and (3) the relation between canonical correlation analysis and chi-square contingency table analysis. The paper shows that canonical correlation analysis can be conceptualized as the statistical model that brings together many other statistical techniques in a unified manner, and the power of this overarching model is significantly increased by applying the concept of partial correlation to the canonical case. Two data sets (one with two Y variables, three X variables, and two classification variables; and the other with two mixed variables with three levels for each) are used to illustrate the points covered. Computer program results are presented to augment the discussion. Appendix A presents the SAS program for some tabulated data. Six tables present analysis results, and there is a 29-item list of references. (Author/SLD)

Canonical Correlation Analysis as a General Analytical Model

Fan, Xitao

Pennsylvania Power & Light Co.

Paper presented at the American Educational Research Association 1992 Annual Meeting (session #28.50), San Francisco, CA, April 22, 1992.

ABSTRACT

This paper focuses on three aspects related to the conceptualization and application of canonical correlation analysis as a dominant statistical model:

1) partial canonical correlation analysis and its application in statistical testing;
2) the relation between canonical correlation analysis and discriminant analysis;
3) the relation between canonical correlation analysis and chi-square contingency table analysis.

The paper shows that canonical correlation analysis can be conceptualized as the statistical model which brings together many other statistical techniques in a unified manner, and the power of this overarching model is significantly increased by applying the concept of partial correlation to the canonical case. Two data sets are used to illustrate the points covered in the paper, and the computer program results are presented to augment the discussion.

The utilization of multivariate statistical analysis has been widely recognized as to be important in social and behavioral science research. The importance stems from the consideration that we intend to honor the complex social reality in which we operate and which we eventually want to generalize to, and we intend to avoid inflating experiment-wise error rate in our statistical analysis (Fish, 1988; Johnson & Wichern, 1988; SAS/STAT User's Guide, Version 6, Vol. 4, 1989; Stevens, 1936). Among the multivariate statistical techniques, canonical correlation analysis has occupied an important strategic position. It has often been conceptualized as a unified approach to almost all parametric statistical testing procedures, univariate and multivariate alike (Baggaley, 1981; Dunteman, 1984; Fornell, 1978; Knapp, 1978; Kshirsagar, 1972; SAS/STAT User's Guide, Version 6, Vol. 4, 1989; Thompson, 1984, 1988, 1991a). It has even been considered to subsume such non-parametric procedure as contingency table analysis (Knapp, 1978; Kshirsagar, 1972).

Many authors have shown, either theoretically or empirically, the equivalence between canonical correlation analysis and many other statistical testing procedures ranging from simple correlation, t-test, to MANOVA, discriminant analysis, and even chi-square contingency table analysis (Knapp, 1971; Kshirsagar, 1972; Tatsuoka, 1989; Thompson, 1988, 1991a). The implications of this intimate relationship between canonical correlation analysis and most of the other statistical testing

procedures are both practically and theoretically meaningful and far-reaching.

Practically, this relationship reveals the fact that researcher's choice of any particular statistical technique in a research situation contributes nothing to the validity of any causal inferences one may choose to make. This point is relevant since, despite repeated warnings of some methodologists (Cook & Campbell, 1971; Thompson, 1981, 1985, 1991b), some unsophisticated researchers still harbor the misconception that OVA methods (ANOVA, ANCOVA, MANOVA, MANCOVA) are more closely related to experimental design, thus permitting valid causal inferences; correlation and regression approaches, on the other hand, are correlational in nature, thus not permitting causal inferences. After all, isn't it true that all of our statistics text books emphasize the fact that correlation does not mean causation? Isn't it also true that Fisher developed and subsequently used OVA methods extensively in his research and made valid causal inferences?

The misconception of relating OVA methods with causal inferences is readily expelled once researchers realize that both OVA methods and regression approaches are the same statistically, and both of them can be considered special cases of canonical correlation analysis. Because of this fact, it is no exaggeration to state that all parametric testing procedures are correlational (Thompson, 1991a, 1991b), and this statement may even extend to non-parametric testing procedure such as chi-square contingency

table (Dunteman, 1984; Knapp, 1971; Kshirsagar, 1972). Viewed from this perspective, no methods permit valid causal inferences without appropriate research design. In reality, what permits valid causal inferences is NOT the statistical techniques we happen to use, but the research design and data collection process. Though Fisher used ANOVA extensively in his research in agriculture, it is his randomization scheme which made the causal inferences in his research valid, not the ANOVA technique he developed and used (Lentner & Bishop, 1986).

Theoretically, the intimate linkage between canonical correlation analysis and other statistical testing procedures shows that, often, the relationship between two groups of variables must be exploited to yield fruitful results in our statistical analysis. As Kshirsagar (1972) explained, "most of the practical problems arising in statistics can be translated, in some form or the other, as the problem of measurement of association between two vector variates X and Y" (p. 281). Since canonical correlation analysis summarizes the relationships between two groups of variables, it therefore brings together OVA methods (both univariate and multivariate), correlation and regression analysis, discriminant analysis, and even chi-square contingency table analysis in a unified manner. The realization of this theoretical unification of statistical techniques under the overarching canonical correlation model elevates our understanding of statistical techniques to a more strategic level, just as the realization that the univariate general linear

model subsumes both univariate OVA methods (ANOVA and ANCOVA) and regression analysis greatly enhances our understanding of the nature of those statistical techniques (Cohen, 1968).

A BRIEF DESCRIPTION OF CANONICAL CORRELATION ANALYSIS

Hotelling (1935) was the first to tackle the problem of identifying and measuring relations between two sets of variables, and he invented and utilized canonical correlation analysis to investigate the relationship between one set of reading variables and the other set of arithmetic variables in a psychology study. Later, this statistical technique was applied to many other research areas.

Canonical correlation analysis can be understood as the bivariate correlation of two synthetic variables which are the linear combinations of the two sets of original variables (Johnson & Wichern, 1988; Thompson, 1984, 1991a). The two sets of original variables are linearly combined to produce pairs of synthetic variables which have maximal correlation, with the restriction that each member of each subsequent set of such synthetic variables is orthogonal to all members of all other sets. The maximum number of such pairs of synthetic variables which can be produced equals the number of variables in the smaller set of the two. In this sense, the synthetic variables in canonical correlation analysis, which are the linear combinations of the original variables, are similar to the synthetic variables

produced in some other multivariate analysis techniques such as principal component analysis, discriminant analysis, etc. The difference, however, is that, in different statistical analysis, the original variables are linearly combined to satisfy different criteria. For example, in principal component analysis, original variables are linearly combined to produce synthetic variables which have maximum variance. In discriminant analysis, the original variables are linearly combined to produce synthetic variables which maximizes the ratio of between-group variance to within-group variance, so that different groups can be maximally differentiated on the synthetic variables.

As can be expected in any multivariate statistical methods, in canonical correlation analysis, eigenstructures of some matrices are involved in deriving the linear coefficients needed to produce the synthetic variables and in deriving the canonical correlation coefficients for different canonical functions. Let us assume that we have two sets of variables as follows and X is the smaller set of the two:

$$X=\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_i \end{bmatrix}, \qquad Y=\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_j \end{bmatrix}$$

When combined, the two sets of variables have the following partitioned variance-covariance matrix[1]:

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

The derivation of the linear coefficients for combining original variables into canonical variates is based on the following two matrices:

$$\begin{aligned} \boldsymbol{A} &= \Sigma_{xx}^{-1}\,\Sigma_{xy}\,\Sigma_{yy}^{-1}\,\Sigma_{yx} \\ \boldsymbol{B} &= \Sigma_{yy}^{-1}\,\Sigma_{yx}\,\Sigma_{xx}^{-1}\,\Sigma_{xy} \end{aligned}$$

The two matrices, **A** and **B**, have the same eigenvalues λ_i, but with different eigenvectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ associated with the eigenvalues λ_i. The elements of the eigenvectors $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, it turns out, are the linear coefficients for the original two sets of variables **X** and **Y** respectively. In this way, we obtain a pair of synthetic variables (canonical variates):

$$\begin{aligned} U_i &= \boldsymbol{a}_i' \boldsymbol{X} = a_1x_1 + a_2x_2 + \ldots + a_ix_i \\ V_i &= \boldsymbol{b}_i' \boldsymbol{Y} = b_1y_1 + b_2y_2 + \ldots + b_jy_j \end{aligned}$$

and the correlation between U_i and V_i is maximized, subject to the restriction that each subsequent canonical function is orthogonal to all previous canonical functions. As a matter of fact, the

[1] Throughout this paper, variance-covariance matrices are used instead of correlation matrices in all mathematical derivations, mainly for the reason that correlation matrix is a special case of variance-covariance matrix, i.e., correlation matrix is just the variance-covariance matrix for standardized variables. Because of this, the variance-covariance matrix is considered to have wider applicability.

correlation between U_i and V_i (R_{UVi}) is the square root of the associated eigenvalue λ_i, i.e., $R^2_{UVi} = \lambda_i$. The maximum number of canonical variate pairs (canonical functions), as stated previously, equals the number of original variables in the smaller set $\mathbf{X}$.

The reasons for utilizing the eigenstructures of matrices $\mathbf{A}$ and $\mathbf{B}$ to derive eigenvalues λ_i ($\lambda_i = R^2_{UVi}$) and eigenvectors a_i and b_i for linear coefficients are mathematical, since the eigenstructures of $\mathbf{A}$ and $\mathbf{B}$ mathematically guarantees the following:

1) correlation between U_i and V_i is maximized to be $\lambda_i^{1/2}$;
2) correlation between U_i and U_j, or U_i and V_j, or V_i and V_j, or V_i and U_j is zero (for any i not equal j), that is, canonical variates across pairs, either within one set or across sets, have zero correlation.
3) $Var(U_i) = Var(V_i) = 1$ (for any i), i.e., canonical variates are standardized to have unit variance.

Since canonical variates are just linear combinations of original variables, the properties of linear combination of random variables apply. Specifically, for our interest, the linear combinations

$$U_i = \mathbf{a}_i' \mathbf{X} = a_1x_1 + a_2x_2 + \ldots + a_ix_i$$
$$V_i = \mathbf{b}_i' \mathbf{Y} = b_1y_1 + b_2y_2 + \ldots + b_jy_j$$

have variances:

$$Var(U_i) = Var(\mathbf{a}_i'\mathbf{X}) = \mathbf{a}_i'\Sigma_{xx}\mathbf{a}_i$$
$$Var(V_i) = Var(\mathbf{b}_i'\mathbf{Y}) = \mathbf{b}_i'\Sigma_{yy}\mathbf{b}_i$$

So the variances of the synthetic canonical variates are easily calculated by using the linear coefficient vectors and variance-covariance matrix of the original variables. Since canonical variates are standardized to have unit variance, we must have:

$$Var(U_i) = a_i' \Sigma_{xx} a_i = 1$$
$$Var(V_i) = b_i' \Sigma_{yy} b_i = 1$$

Several authors have amply demonstrated, either theoretically or empirically, the equivalence of canonical correlation analysis with almost all other parametric statistical testing procedures (Baggaley, 1981; Dunteman, 1984; Knapp, 1978; Kshirsagar, 1972; Tatsuoka, 1989; Thompson, 1988, 1991a; Zinkgraf, 1983). So it is not the purpose of this paper to provide comprehensive coverage on this interesting topic, and readers are referred to the sources above for more detailed and comprehensive account on this topic. Instead, three aspects of canonical correlation analysis, when conceptualized as an overarching model for statistical analysis, will be discussed in some detail in the following sections. The three aspects are:

1) partial canonical correlation analysis and its application in statistical testing;
2) canonical correlation approach to discriminant analysis; and
3) canonical correlation approach to contingency table analysis.

I. PARTIAL CANONICAL CORRELATION ANALYSIS AND ITS APPLICATION IN STATISTICAL TESTING

Description of Partial Canonical Correlation Analysis

Partial canonical correlation is the natural generalization of partial correlation from the univariate to the multivariate situation. In univariate situation, the need for partial correlation arises when we have, e.g., three variables x, y and z, and they are intercorrelated with each other. If we are interested in removing the effect of z on both x and y, then determining the relationship between x and y, what we have is the partial correlation between x and y after removing z's influence on both. Mathematically, the situation described above is the simplest partial correlation we can encounter, and the partial correlation between x and y after partialing out z's influence can be expressed mathematically as follows (Glass & Hopkins, 1984; Neter, Wasserman & Kutner, 1989):

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Simple partial correlation (as opposed to partial canonical correlation to be discussed later) has some practical implications which are not often realized by some researchers. Glass and Hopkins (1984) offer a good example. Suppose we are examining the relationship between reading performance and visual perceptual ability of children within certain range of age, we may have high positive relationship which indicates that those with high visual perceptual ability also tend to have high

reading performance, and vice varsa. But this may be due to the fact that both kinds of ability are developmental, thus both are positively related to age. Once we hold age constant, the observed high positive relationship between reading performance and visual perceptual ability may drastically decrease or even disappear. If this scenario is true, the partial correlation between reading performance and visual perceptual ability after partialing out the effect of age on both will give us the indication of the true relationship between the two kinds of ability, while regular bivariate Pearson correlation may be erroneously misleading (Glass & Hopkins, 1984). The derivation of simple partial correlation coefficient as well as the testing for its statistical significance is readily available through major statistical software packages such as SAS and SPSS.

The concept of simple partial correlation can be extended to canonical correlation analysis. The mathematical foundation for this extension is offered in Anderson (1984), Cooley and Lohnes (1971), Johnson and Wichern (1988) and Timm (1975). An excellent substantive research example employing partial canonical correlation analysis is offered by Cooley and Lohnes (1971). They investigated, among other things, the canonical correlational relationship between two sets of variables measuring Grade 12 abilities and Grade 9 interests after partialing out the effect of another set of variables which measures Grade 9 abilities. Interested readers are referred to Cooley and Lohens (1971) for the detailed description of the research example and some

practical interpretation of the partial canonical correlation analysis.

The situation demanding partial canonical correlation analysis arises when we have three vectors of random variables

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_i \end{bmatrix}, \qquad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_j \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_k \end{bmatrix}$$

and these three vectors of random variables are intercorrelated with each other. As in the case of simple partial correlation, we may be interested in investigating the relationship between vector **X** and vector **Y** after partialing out the effect of vector **Z** on both **X** and **Y** vectors. Conceptually, this is a problem of partial correlation, but multivariate vectors are involved instead of simple univariate variables as in the case of simple partial correlation.

The crucial point in performing partial canonical correlation analysis is to find out from what variance-covariance matrices partial canonical correlation coefficients, as well as linear coefficients for combining original variables into canonical variates, can be derived. It turns out that the variance-covariance matrices of conditional distribution of **X** and **Y**, given **Z**, i.e., the residualized variance-covariance matrices of **X** and **Y** after partialing out the effect of vector **Z** from both **X** and **Y**, provide us with the solutions. More specifically,

suppose that, when combined, the three vectors of variables, X, Y and Z have the following partitioned variance-covariance matrix :

$$\Sigma_{XYZ} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}$$

The conditional variance-covariance matrix of X and Y, given Z, that is, the residualized variance-covariance matrix of X and Y after partialing out the effect of Z, is given as (Anderson, 1984; Johnson & Wichern, 1988; Timm, 1975):

$$\Sigma_{XY.Z} = \begin{bmatrix} \Sigma_{xx.z} & \Sigma_{xy.z} \\ \Sigma_{yx.z} & \Sigma_{yy.z} \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{xy} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zy} \\ \Sigma_{yx} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zx} & \Sigma_{yy} - \Sigma_{yz}\Sigma_{zz}^{-1}\Sigma_{zy} \end{bmatrix}$$

Similar to regular canonical correlation analysis, the eigenvalues λ_i from the following two matrices, A_{part} and B_{part}, will be the squared partial canonical correlation coefficients for ith canonical functions, and the eigenvectors a_i and b_i associated with the eigenvalue λ_i will be the linear coefficient vectors (function coefficients) which combine the original variables into synthetic canonical variates.

$$A_{part} = \Sigma_{xx.z}^{-1}\Sigma_{xy.z}\Sigma_{yy.z}^{-1}\Sigma_{yx.z}$$
$$B_{part} = \Sigma_{yy.z}^{-1}\Sigma_{yx.z}\Sigma_{xx.z}^{-1}\Sigma_{xy.z}$$

Again, as in regular canonical correlation analysis, the two matrices, A_{part} and B_{part}, have the same eigenvalues but with different eigenvectors. More detailed mathematical explanation of

conditional distribution of vectors of random variables and the derivation of partial canonical correlation functions can be found in Anderson (1984), Johnson and Wichern (1988) and Timm (1975).

As with regular canonical correlation functions, partial canonical correlation functions can be tested for statistical significance, and the testing procedure is similar to testing regular canonical correlation functions for both overall test and sequential test. Timm (1975) offers the necessary details for testing partial canonical correlation functions for their statistical significance.

Application of Partial Canonical Correlation in Statistical Testing

Partial canonical correlation analysis is not only applicable in substantive research, as Cooley and Lohnes' excellent research example has shown (Cooley & Lohnes, 1971), it can also be used as a convenient testing procedure when canonical correlation analysis is employed as the overarching model which subsumes other parametric testing procedures.

It is well documented that by adopting some artificial coding scheme such as "dummy" coding or contrast coding to represent group membership, all OVA methods (ANOVA, MANOVA, ANCOVA, MANCOVA) can be translated into canonical correlation analysis problem (Knapp, 1978; Kshirsagar, 1972; Thompson, 1988, 1991a). Although adopting the canonical correlation analysis approach to OVA methods is straightforward for an omnibus test,

conventionally, subsequent testing for individual factors and factor interactions is procedurally tedious. This is so mainly because the Wilk's Λ from the reduced model (correlating dependent variables with independent dummy variables after dropping those associated with the factor or interaction of factors to be tested) CANNOT be used directly for the purpose of testing; instead, the ratio of full model Λ to reduced model Λ will be calculated, and this ratio becomes the Wilk's Λ for the factor or interaction of factors to be tested. Rao's F approximation is then applied to the ratio and regular testing for statistical significance can be carried out for the effect of a particular factor or interaction of factors. More detailed description of this transformation procedure is provided in Thompson (1988, 1991a) and Zinkgraf (1983).

The problem of testing for factor or interaction effects when using a canonical correlation approach to OVA methods can be translated into partial canonical correlation analysis problem. Since partial canonical correlation functions can be directly tested for statistical significance, the approach provides us a convenient means for testing factor or interaction effect. It can be recalled that in regular OVA methods, to test for factor or interaction effect is to test for statistical significance of the marginal contribution of the factor or interaction accounting for the variance (covariance matrix, in multivariate case) of the dependent variable(s), given that all other factor(s) are already in the model. This is equivalent, if we use canonical correlation

approach to OVA methods, to testing for partial canonical correlation functions between dependent variables and those dummy variable(s) representing the factor of interest, after the effects of all other independent variables representing other factors or interactions have been partialed out from the model. Using partial canonical correlation analysis, the effect of any factor or interaction can be tested directly by partialing out the effects of other independent variables, thus the conventionally required transformation of Wilk' Λs, as illustrated in Thompson (1988, 1991a) and described in Zinkgraf (1983), for testing factor or interaction effect can be avoided.

Table 1 presents a small data set which will be used to illustrate some points explained in this paper. The data set has two Y variables, three X variables, two classification variables A and B. The two classification variables are also represented by contrast coding A1 (for the two levels of A factor) and B1 and B2 (for the three levels of B factor). The interaction between the two classification variables are represented by AB1 and AB2, which are the multiplication of A1 by B1 and A1 by B2, respectively. For detailed explanation of using coding scheme to represent classification variables, readers are referred to Cohen (1968), Kerlinger and Pedhazur (1973), Ott (1988), and Neter, Wasserman and Kutner (1989).

Insert Table 1 About Here

Among the major statistical software packages, currently SAS is the only one which performs partial canonical correlation analysis as a standard option under canonical correlation analysis procedure PROC CANCORR. Cooley and Lohnes (1971) offers a computer program for performing partial canonical correlation analysis also. Appendix A presents a SAS program which performs the various kinds of statistical analysis to be illustrated in the paper.

The equivalence of MANOVA with partial canonical correlation analysis is demonstrated in Table 2. In canonical correlation approach to MANOVA, tests for individual factors and their interaction are accomplished by sequentially performing three partial canonical correlation analyses between dependent variables and those dummy variables representing the factor to be tested, while partialing out the effects of all other dummy variables not being tested. This partialing, in fact, is testing the significance of additional contribution of a particular factor, given that the effects of the other factor and the interaction have already been taken into consideration. Conceptually, this is similar to the full model vs. reduced model approach illustrated in Thompson (1988, 1991a) and explained in Zingkraf (1983), but by invoking the concept of partial canonical correlation, the test for individual factors and interactions in MANOVA is translated into direct test of partial canonical functions. In this way, transformations of Wilk's Λs from the

full and reduced models, and the subsequent calculation of Rao's F approximation, is avoided here.

Insert Table 2 About Here

The partial canonical correlation analysis described above can be applied to any OVA methods for testing individual factors, factor interaction, as well as covariate(s). For each factor (or interaction, or covariate), a separate partial canonical correlation analysis is performed between dependent variable(s) and those independent variables (dummy for classification variables, continuous for covariate) representing the factor (or interaction, or covariate) to be tested, while partialing out the effects of all other independent variables for other factors and interactions. (Interested readers are encouraged to use one dependent variable in order to see the equivalence in the case of ANOVA.)

Not only can partial canonical correlation analysis be applied to statistical testing in OVA methods, it can also be conveniently used to solve some problems in multiple regression analysis. In multiple regression analysis, very often, we try to find the best set of predictors for our dependent variable. In trying to do so, we sometimes encounter the situation in which we have to determine what the effect will be if several independent variables are added to the model simultaneously. In other words, we have to assess the marginal contribution of two or more

independent variables as a group to the variance of the dependent variable, given that some other independent variables are already in the model. Conventionally, testing the additional effect of two or more variables as a group in regression analysis is not direct, to say the least; and it is cumbersome in some sense. Even major statistical packages such as SAS or SPSS do not provide readily available results. Competing models (with and without the added group of independent variables) have to be run, and either sums of squares or coefficients of determination (multiple R^2s) from the two competing models must be used for testing the significance of the marginal contribution of the added group of variables in the regression model. The formulas for using sums of squares and coefficients of determination from the two competing models (full and reduced) are as follows (Glass & Hopkins, 1984; Neter, Wasserman & Kutner, 1989):

$$F_{v1,v2} = \frac{(R_F^2 - R_R^2)/(df_R - df_F)}{(1 - R_F^2)/df_F}$$

$$F_{v1,v2} = \frac{(SSE_R - SSE_F)/(df_R - df_F)}{SSE_F/df_F}$$

In these equations, SSE represents Sum of Squares due to error from respective models, and v1 = $df_R - df_F$, v2 = df_F.

The problem of testing for significance of the marginal contribution of a new group of independent variables in regression model is easily solved by adopting the partial canonical correlation analysis approach. Since we want to assess the marginal contribution of a group of new variables, given that

other independent variables are already in the regression model, we simply perform a partial canonical correlation analysis between the dependent variable and the group of new independent variables, while partialing out the effects of other independent variables already in the model. Testing the significance of the partial canonical function is equivalent to testing the significance of the marginal contribution of the added group of variables in the multiple regression model.

Table 3 presents the results of two competing regression models, Full (2) vs. Reduced (1) models, to test the significance of the marginal contribution of X2 and X3, given that X1 is already in the model. Results from partial canonical correlation approach are presented for comparison.

$$(1) \qquad Y_1 = \beta_0 + \beta_1 X_1 + \epsilon$$
$$(2) \qquad Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

The F test for the significance of the marginal contribution of X2 and X3 as a group is illustrated using both formulas presented earlier. Thus it is seen that the significance testing for partial canonical correlation function is exactly the same as the significance testing results from using either of the two formulas.

Insert Table 3 About Here

II. CANONICAL CORRELATION APPROACH TO DISCRIMINANT ANALYSIS

The relationship between canonical correlation analysis and discriminant analysis has been demonstrated by several authors

(Dunteman, 1984; Kshirsagar, 1972; Tatsuoka, 1989; Thompson, 1984, 1991a). The basic procedure is, instead of doing discriminant analysis, dummy coding is constructed to represent group membership, and canonical correlation analysis is performed between the random predictor variables and the dummy coding variables for group membership. The results from both approaches are the same, with the exception that the function coefficients from the two approaches are not the same in terms of numeric value, instead, they are the same in terms of ratio. This can be seen if we rescale the two vectors of coefficients from the two approaches so that the largest element in each set equals one, as Tatsuoka (1989) demonstrated. Table 4 presents the results from the two approaches, discriminant analysis and canonical correlation approach to discriminant analysis. In discriminant analysis, the classification variable is A, and two X variables (X1 and X2) are predictor variables. In the canonical correlation approach, canonical correlation analysis is performed between two X variables (X1 and X2) and A1, which is the dummy coding to represent the two levels in A.

Since the function coefficients from the two approaches are not the same numerically, naturally, the question arises as to whether the synthetic variables constructed from the two different sets of coefficients can equally differentiate the groups. After all, to construct synthetic variables which maximally discriminate the groups is the major purpose of discriminant analysis (Huberty & Wisenbaker, 1991; Johnson &

Wichern, 1988;). The answer to the above question is a definite yes. Mathematically, the two sets of coefficients from the two approaches are the same in the sense that when used to combine original variables into synthetic variables (discriminant functions), the groups are equally differentiated to the maximum degree no matter which set we happen to use. The reasons for regarding these two as equivalent can be furnished from several angles.

First of all, geometrically, the two sets of coefficients from the two approaches are two vectors in multi-dimensional space. We are only concerned with the direction of the vectors, not the magnitude of them, since these vectors are arbitrarily scaled. As long as the elements in the vectors are proportional, they are in the same direction, and they can always be rescaled to be equal in numeric value by multiplication with a constant. As Kshirsagar observed (1972), unlike multiple regression where the coefficients are unique, in discriminant analysis (in canonical correlation analysis, too, for this matter), the coefficients are not unique in numeric value; they are only unique in their ratio. So theoretically, for one discriminant function, there are an infinite number of coefficient sets which can do an equally good job in differentiating the groups. The same is true with canonical correlation analysis: canonical correlation coefficient will not be affected by multiplying the coefficient vectors with any constant.

In discriminant analysis, the condition for maximal differentiation of the groups on the synthetic variables is that the ratio of between-group variance to pooled within-group variance on the synthetic variable (discriminant function) is maximized (Johnson & Wichern, 1988; Kshirsagar, 1972), subject to the restriction that each discriminant function is orthogonal to previous ones. Let U_i be the ith discriminant function, and $\mathbf{a}_i$ the ith function coefficient vector, and we have the linear combination of the original variables:

$$U_i = \mathbf{a}_i'\mathbf{X} = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

Let $\mathbf{B}_0$ be the between-group variance-covariance matrix, and $\mathbf{S}_{pooled}$ be the pooled within-group variance-covariance matrix of the original $\mathbf{X}$ variables (X1 and X2) respectively. The between-group variance and the pooled within-group variance on the synthetic variable, using the properties of linear combination of random variables, are

$$Var\ (U_i)_B = \mathbf{a}_i'\mathbf{B}_0\mathbf{a}_i$$
$$Var\ (U_i)_{Pool.W} = \mathbf{a}_i'\mathbf{S}_{pooled}\mathbf{a}_i$$

and the ratio

$$\frac{Var\ (U_i)_B}{Var\ (U_i)_{Pool.W}} = \frac{\mathbf{a}_i'\mathbf{B}_0\mathbf{a}_i}{\mathbf{a}_i'\mathbf{S}_{pooled}\mathbf{a}_i}$$

is maximized in discriminant analysis.

In our example of discriminant analysis, we have the following variance-covariance matrices for the original two X variables, X1 and X2:

$$B_0 = \begin{bmatrix} 12.0868 & 11.2674 \\ 11.2674 & 10.5035 \end{bmatrix}, \quad S_{pooled} = \begin{bmatrix} 5.1326 & 2.4129 \\ 2.4129 & 4.9962 \end{bmatrix}$$

and the first discriminant function (since we have only two groups, this is the only discriminant function we can obtain) has the coefficient vector (raw coefficients):

$$a_D = \begin{bmatrix} 0.2771 \\ 0.2392 \end{bmatrix}$$

Thus, we have the ratio of between-group variance to pooled within-group variance on the discriminant function as:

$$\begin{aligned}\frac{Var\ (U_1)_B}{Var\ (U_1)_{Pool.W}} &= \frac{a_D' B_0 a_D}{a_D' S_{pooled} a_D} \\ &= \frac{[0.2771\ 0.2392]\begin{bmatrix} 12.0868 & 11.2674 \\ 11.2674 & 10.5035 \end{bmatrix}\begin{bmatrix} 0.2771 \\ 0.2392 \end{bmatrix}}{[0.2771\ 0.2392]\begin{bmatrix} 5.1326 & 2.4129 \\ 2.4129 & 4.9962 \end{bmatrix}\begin{bmatrix} 0.2771 \\ 0.2392 \end{bmatrix}} \\ &= \frac{3.023}{1} = 3.023\end{aligned}$$

If we substitute the coefficient vector from canonical correlation approach for that from the discriminant analysis, the ratio of the between-group variance to the pooled within-group variance on the synthetic variable will be exactly the same as in discriminant analysis. This is easily verified by using the coefficient vector from our canonical correlation approach:

$$\frac{\mathbf{a}_C'\mathbf{B}_0\mathbf{a}_C}{\mathbf{a}_C'\mathbf{S}_{pooled}\mathbf{a}_C} = \frac{[0.1741\ \ 0.1503]\begin{bmatrix}12.0868 & 11.2674\\ 11.2674 & 10.5035\end{bmatrix}\begin{bmatrix}0.1741\\ 0.1503\end{bmatrix}}{[0.1741\ \ 0.1503]\begin{bmatrix}5.1326 & 2.4129\\ 2.4129 & 4.9962\end{bmatrix}\begin{bmatrix}0.1741\\ 0.1503\end{bmatrix}}$$

$$= \frac{1.1931}{0.3947} = 3.023$$

The equivalence of the two ratios proves the fact that the synthetic variable constructed from either of the coefficient vectors from the two approaches discriminate the groups equally well, thus they are considered equivalent, despite the superficial numeric difference.

The reason for the numeric difference of the coefficient vectors from the two approaches is arbitrary scaling. Conventionally, canonical correlation functions are scaled to have unit variance, while discriminant functions are scaled so that the POOLED within-group variance equals one. This means that for separate groups, the discriminant functions may not have unit variance; neither may the discriminant functions have unit variance when groups are combined. But when the variances for separate groups are pooled, the pooled variance of the discriminant function satisfies the condition:

$$Var(U_i)_{pooled} = 1$$

Since

$$Var(U_i)_{pooled} = \mathbf{a}_D'\mathbf{S}_{pooled}\mathbf{a}_D$$

we must have

$$\mathbf{a}_D'S_{pooled}\mathbf{a}_D = 1$$

and the previous calculation on the between-group vs. pooled within-group ratio has already verified this.

Since the coefficient vectors from the two approaches are arbitrarily scaled to satisfy different criteria, naturally they become different in numeric value, though they are still equal in terms of differentiating groups. To convert the coefficient vector from canonical correlation approach so that it equals that of discriminant analysis in numeric value is a problem of rescaling. Without this rescaling, the synthetic variable constructed using the coefficients from canonical correlation approach will not have the property of unit pooled within-group variance, as our results shows:

$$Var(U_1)_{pooled} = \mathbf{a}_C'S_{pooled}\mathbf{a}_C$$
$$= [0.1741\ \ 0.1503]\begin{bmatrix}5.1326 & 2.4129\\2.4129 & 4.9962\end{bmatrix}\begin{bmatrix}0.1741\\0.1503\end{bmatrix} = 0.3947$$

To rescale the coefficient vector from canonical correlation approach so that it satisfies the conventional condition of discriminant analysis, we only need to standardize the coefficient vector in a way similar to any other standardization process, i.e., to divide each element in the coefficient vector by some sort of standard deviation. We recall that if we use $\mathbf{a}_C$, the coefficient vector from canonical correlation approach, to construct synthetic variable as discriminant function, the corresponding discriminant function has the pooled within-group variance $\mathbf{a}_C'S_{pooled}\mathbf{a}_C$, and which does not equal 1, as demonstrated

previously. The square root of $a_C' S_{pooled} a_C$ is the pooled within-group standard deviation on such a discriminant function, and it is this standard deviation that we can use to standardize the coefficient vector from canonical correlation approach. More specifically, if a_D is the coefficient vector from discriminant analysis, and a_C is its counterpart from canonical correlation approach, S_{pooled} is the pooled within-group covariance matrix of the original predictor variables, the two coefficient vectors are related as follows:

$$a_D = \frac{a_C}{\sqrt{a_C' S_{pooled} a_C}}$$

Earlier, we have calculated the $a_C' S_{pooled} a_C$ to be 0.3947. Using the coefficient vectors from the two approaches, we easily verify

$$\begin{aligned} \frac{a_C}{\sqrt{a_C' S_{pooled} a_C}} &= \frac{1}{\sqrt{0.3947}}\begin{bmatrix}0.1741\\0.1503\end{bmatrix} \\ &= \frac{1}{0.6282}\begin{bmatrix}0.1741\\0.1503\end{bmatrix} = \begin{bmatrix}0.2771\\0.2392\end{bmatrix} = a_D \end{aligned}$$

The conversion of standardized coefficients from the two approaches is the same, and only pooled covariance matrix (S_{pooled}) from standardized predictor variables needs to be used.

The rescaling is conceptually straightforward, but procedurally tedious, since the pooled covariance matrix of the original variables has to be found, and matrix calculations are involved. At this point, it should be emphasized that this rescaling is NOT a mathematical necessity, but rather, it is done simply to satisfy certain arbitrary conventional condition.

III. CANONICAL CORRELATION APPROACH TO CONTINGENCY TABLE ANALYSIS

Several authors have discussed the relationship between canonical correlation analysis with the contingency table analysis (Dunteman, 1984; Knapp, 1978; Kshirsagar, 1972). In some sense, this is an extreme case of stretching the application of canonical correlation analysis (Knapp, 1978), since, for this application, all the variables in canonical correlation analysis are "dummy" variables representing group membership, and none of them is a true random variable.

In contingency table analysis, we have R levels (R rows) on one classification variable, and C levels (c columns) on the other classification variable. According the subject's level on the two nominal variables, the subject is entered into the appropriate cells in the contingency table. As a result, we have a contingency table with R rows and C columns. The observed frequency and expected frequency for each cell are used for calculation of χ^2 statistic which is subsequently used for testing the hypothesis of independence of row and column classification probabilities.

The procedure for adopting canonical correlation approach to contingency table analysis is similar to other situations. For R rows and C columns, we use r-1 and c-1 dummy coded variables to represent the levels on the two classification variables So for each subject, r-1 plus c-1 dummy variables will be used to represent its position in the contingency table. Finally, a canonical correlation analysis is to be performed between r-1 and

c-1 dummy variables, and testing for the canonical correlation functions from this approach is equivalent to that based on the classical chi-square test for independence.

Table 5 presents part of a data set with two nominal variables, A and B, with three levels for each. For canonical correlation approach, we need two dummy variables (A1, A2) to represent Variable A and another two dummy variables (B1, B2) for Variable B. All these are included in Table 5.

Insert Table 5 About Here

Table 6 presents the results of both classical chi-square test and canonical correlation approach to contingency table analysis. From the canonical correlation approach, the probability from Wilk's Λ deviates a little from the probability from chi-square test (0.001 vs. 0.0009), but the probability derived from Pillai's Trace does equal that from chi-square test. The reason for the small discrepancy is related to sample size, since χ^2 statistic is related to F statistic as follows (Knapp, 1978):

$$\chi^2_{(r-1)(c-1)} = (r-1)(c-1)F_{(r-1)(c-1),\infty}$$

If sample size increases, the probabilities from the two approaches will converge. When sample size is large enough, χ^2 statistic can be directly calculated from F in canonical approach by multiplying the canonical F by (r-1)(c-1). For our example, the sample size is not large enough, and the results from the two approaches have not quite converged:

$$(r-1)(c-1)F = 4 \times 4.7476 = 18.9904 > \chi^2 = 18.194$$

Another more interesting fact about these two approaches is the relationship between χ^2 statistic and Pillai's trace from canonical correlation analysis. As Kshirsagar proved mathematically (Kshirsagar, 1972, p. 383), the two are related as (N: total sample size):

$$\text{Pillai's Trace } V = \frac{\chi^2}{N}$$

This relationship is easily verified using our example results:

$$N(\text{Pillai's trace}) = 200 \times 0.09097222 = 18.19444 = \chi^2$$

Because of this relationship, classical χ^2 test is considered to be the same as the test based on Pillai's trace (Kshirsagar, 1972), even if when sample size is only moderate.

This application of canonical correlation analysis indeed stretches it to its limits, since almost no assumption exists for the nominal data in the contingency table, not to mention multivariate normality. Viewed in this perspective, the little discrepancy between the two approaches when the sample size is not adequately large should be quite tolerable.

CONCLUSION

Canonical correlation analysis is a powerful overarching statistical model which brings together many statistical techniques in a unified manner. Understanding the relationship between canonical correlation analysis and other statistical techniques is important for any of us who are ambitious enough as

to try to be seasoned researchers, since this will certainly enhance our grasp of statistical methods, thus elevating us to a higher and more strategic position in the hierarchy of statistical techniques. Under the framework of canonical correlation analysis, extending the concept of partial correlation to canonical case has some interesting positive implications. This extension makes canonical correlation analysis more applicable as a statistical testing tool, thus increasing its power as a dominant statistical model. Furthermore, the application of canonical correlation analysis as a general data analytic system will most likely be facilitated in research practice by this extension.

REFERENCES

Anderson, T.W. (1984). An introduction to multivariate statistical analysis (2nd ed.). New York: Wiley.

Baggaley, A.R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5(1), 123-131.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443,

Cook, T.D., & Campbell, D.T. (1971). The design and analysis of quasi-experiments for field settings. Chicago: Rand McNally.

Cooley, W.W., & Lohnes, P.R. (1971). Multivariate data analysis. New York: Wiley.

Dunteman, G.H. (1984). Introduction to multivariate analysis. Beverly Hills, CA: Sage Publications, Inc.

Fish, L.J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Fornell, C. (1978). Three approaches to canonical analysis. Journal of the Market Research Society, 20, 166-181.

Glass, G.V., & Hopkins, K.D. (1984). Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice Hall.

Hotelling, H. (1935). The most predictable criterion. Journal of Educational Psychology, 26, 139-142.

Huberty, C.J. & Wisenbaker, J.M. (in press). Discriminant analysis: Potential improvements in typical practice. In B.

Thompson (Ed.), <u>Advances in social science methodology</u> (Vol. 2, pp. 169-209). Greenwich: JAI Press Inc.

Johnson, R.A. & Wichern, D.W. (1988). <u>Applied multivariate statistical analysis</u> (2nd Ed.). Englewood Cliffs, NJ: Prentice Hall.

Kerlinger, F.N. & Pedhazur, E.J. (1973). <u>Multiple regression in behavioral research</u>. New York: Holt, Rinehart and Winston.

Kshirsagar, A.M. (1972). <u>Multivariate analysis</u>. New York: Marcel Dekker, Inc.

Knapp, T.R. (1978). Canonical correlation analysis: A general parametric significance testing system. <u>Psychological Bulletin</u>, <u>85</u>, 410-416.

Lentner, M., & Bishop, T. (1986). <u>Experimental design and analysis</u>. Blacksburg, VA: Valley Book Company.

Neter, J., Wasserman, W. & Kutner, M.H. (1989). <u>Applied linear regression models</u> (2nd Ed.). Boston, MA: Irwin Inc.

Ott, L. (1988). <u>An introduction to statistical methods and data analysis</u> (3rd Ed.). Boston, MA: PWS-KENT Publishing Company.

SAS Institute, Inc. (1989). <u>SAS/STAT user's guide, Version 6, 4th ed, Vol. 1</u>. Cary, NC: SAS Institute Inc.

Stevens. J. (1984). <u>Applied multivariate statistics for the social sciences</u>. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Tatsuoka, M.M. (1989). <u>Multivariate analysis: Techniques for educational and psychological research</u> (2nd ed.). New York: Wiley.

Thompson, B. (1981, November). The problem of OVAism. Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington. (Order document #03980 from National Auxiliary Publication Service, P.O. Box 3513, Grand Central Station, NY, NY 10017)

Thompson, B. (1984). Canonical correlation analysis: Use and interpretation. Sage University Paper series on Quantitative Application in Social Sciences, series no. 07-047. Beverly Hills and London: Sage Pubns.

Thompson, B. (1985). Alternate methods for analyzing data from experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1988). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1991a). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.

Thompson, B. (1991b). Review of Data analysis for research designs by G. Keppel & S. Zedeck. Educational and Psychological Measurement, 51, 500-510.

Timm, N.H. (1975). Multivariate analysis with application in education and psychology. Monterey, CA: Brooks/Cole.

Zinkgraf, S.A. (1983). Performing factorial multivariate analysis of variance using canonical correlation analysis. Educational and Psychological Measurement, 43, 63-68.

Appendix A

SAS Program I: Program for Table 1 Data

```
DATA D1; INFILE AAA;
INPUT ID Y1 Y2 X1 X2 X3 A B A1 B1 B2 AB1 AB2;
  TITLE1 'MANOVA & CANONICAL CORRELATION APPROACH TO MANOVA';
PROC GLM;
  TITLE2 'MANOVA TO TEST FACTOR A, B & A*B INTERACTION';
     CLASS A B;
     MODEL Y1 Y2=A B A*B/NOUNI;
     MANOVA H=_ALL_/SUMMARY;
PROC CANCORR SHORT;
  TITLE2 'PARTIAL CANCORR APPROACH TO TEST A*B INTERACTION';
     VAR Y1 Y2;
     WITH    A1 B1 B2 AB1 AB2;
     PARTIAL A1 B1 B2;
PROC CANCORR SHORT;
  TITLE2 'PARTIAL CANCORR APPROACH TO TEST FACTOR A EFFECT';
     VAR Y1 Y2;
     WITH    A1 B1 B2 AB1 AB2;
     PARTIAL    B1 B2 AB1 AB2;
PROC CANCORR SHORT;
  TITLE2 'PARTIAL CANCORR APPROACH TO TEST FACTOR B EFFECT';
     VAR Y1 Y2;
     WITH    A1 B1 B2 AB1 AB2;
     PARTIAL A1       AB1 AB2;
RUN;
*;
*;
  TITLE1 'PARTIAL CANCORR TO SOLVE MULTIPLE REGRESSION PROBLEM';
PROC REG;
  TITLE2 'MULTIPLE REGRESSION, REDUCED MODEL';
     MODEL Y1=X1;
PROC REG;
  TITLE2 'MULTIPLE REGRESSION, FULL MODEL';
     MODEL Y1=X1 X2 X3;
PROC CANCORR SHORT;
  TITLE2 'PARTIAL CANCORR TO TEST SIGNIFICANCE OF X2 & X3';
     VAR Y1;
     WITH    X1 X2 X3;
     PARTIAL X1;
RUN;
*;
*;
 TITLE1 'CANCORR APPROACH TO DISCRIMINANT ANALYSIS';
PROC CANDISC BCOV PCOV PCORR;
  TITLE2 'CANDISC PROCEDURE FOR STANDARDIZED VARIABLES';
     CLASS A;
     VAR ZX1 ZX2;
PROC CANDISC BCOV PCOV PCORR;
  TITLE2 'CANDISC PROCEDURE FOR ORIGINAL VARIABLES';
```

```
     CLASS A;
     VAR X1 X2;
PROC CANCORR ALL;
  TITLE2 'CANCORR APPROACH TO DISCRIMINANT ANALYSIS';
     VAR   X1 X2:
     WITH  A1;
```

SAS Program II: Program for Table 5 Data

```
DATA D2; INFILE BP3;
INPUT A B A1 A2 B1 B2;
   TITLE1 'CANCORR APPROACH TO CONTINGENCY TABLE';
PROC FREQ;
  TITLE2 'CHI-SQUARE TEST OF CONTINGENCY TABLE ANALYSIS';
     TABLE A*B/CHISQ EXPECTED NOPERCENT NOROW NOCOL;
PROC CANCORR SHORT;
  TITLE2 'CANCORR APPROACH TO CONTINGENCY TABLE ANALYSIS';
    VAR  A1 A2;
    WITH B1 B2;
```

Table 1: Data Set 1

ID	Y1	Y2	X1	X2	X3	A	B	A1	B1	B2	AB1	AB2
1	93	96	9	12	20	1	1	1	1	-1	1	-1
2	88	91	7	10	15	1	2	1	0	2	0	2
3	95	100	8	12	26	1	3	1	-1	-1	-1	-1
4	95	97	10	14	21	1	1	1	1	-1	1	-1
5	95	99	9	12	25	1	2	1	0	2	0	2
6	99	111	10	18	31	1	3	1	-1	-1	-1	-1
7	99	105	8	10	34	1	1	1	1	-1	1	-1
8	81	93	7	9	16	1	2	1	0	2	0	2
9	95	104	5	14	30	1	3	1	-1	-1	-1	-1
10	88	95	10	12	15	1	1	1	1	-1	1	-1
11	99	115	5	11	42	1	2	1	0	2	0	2
12	87	92	9	9	16	1	3	1	-1	-1	-1	-1
13	101	103	13	14	29	2	1	-1	1	-1	-1	1
14	102	107	10	15	32	2	2	-1	0	2	0	-2
15	110	122	18	20	51	2	3	-1	-1	-1	1	1
16	102	108	10	17	31	2	1	-1	1	-1	-1	1
17	106	120	14	18	39	2	2	-1	0	2	0	-2
18	103	109	12	17	32	2	3	-1	-1	-1	1	1
19	103	112	16	17	34	2	1	-1	1	-1	-1	1
20	103	110	11	14	35	2	2	-1	0	2	0	-2
21	105	114	12	15	37	2	3	-1	-1	-1	1	1
22	107	121	16	19	39	2	1	-1	1	-1	-1	1
23	106	118	14	16	39	2	2	-1	0	2	0	-2
24	106	120	10	16	49	2	3	-1	-1	-1	1	1

Data adapted from R.A. Johnson & D.W. Wichern. (1988). Applied Multivariate Statistical Analysis (2nd Ed.), Exercise 9.16, p. 435.

Table 2: Results of MANOVA & Partial Canonical Correlation Approach to MANOVA

1. MANOVA Results

Manova Test for Factor A Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.31290643	18.6647	2	17	0.0001

Manova Test for Factor B Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.80570847	0.9696	4	34	0.4369

Manova Test for A*B Interaction Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.91174914	0.4019	4	34	0.8059

2. Partial Canonical Correlation Results

Partial Canonical Correlation to Test Factor A Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.31290643	18.6647	2	17	0.0001

Partial Canonical Correlation to Test Factor B Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.80570847	0.9696	4	34	0.4369

Partial Canonical Correlation to Test A*B Interaction Effect

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.91174914	0.4019	4	34	0.8059

Table 3: Results of Using Partial Canonical Correlation to Solve Multiple Regression Problem

1. Multiple Regression Results

Model: Reduced Model
Predicted: Y1; Predictor: X1

Analysis of Variance

Source	DF	SS	MS	F Value	Prob>F
Model	1	581.59436	581.59436	19.277	0.0002
Error	22	663.73898	30.16995		
Total	23	1245.33333			

R-square 0.4670

Model: Full Model
Predicted: Y1; Predictor: X1, X2, X3

Analysis of Variance

Source	DF	SS	MS	F Value	Prob>F
Model	3	1131.47524	377.15841	66.251	0.0001
Error	20	113.85809	5.69290		
Total	23	1245.33333			

R-square 0.9086

$$F_{2,20}=\frac{(R_F^2-R_R^2)/(df_R-df_F)}{(1-R_F^2)/df_F}=\frac{(.9086-.467)/2}{(1-.9086)/20}=48.30 \qquad (p < .0001)$$

$$F_{2,20}=\frac{(SSE_R-SSE_F)/(df_R-df_F)}{SSE_F/df_F}=\frac{(663.74-113.86)/2}{113.86/20}=48.29 \qquad (p<.0001)$$

2. Partial Canonical Correlation Analysis Results

Testing Marginal Contribution of X2 and X3

Multivariate Statistics and Exact F Statistics

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.17154047	48.2953	2	20	0.0001

Table 4: Results from Canonical Analysis and Discriminant Analysis

1. Canonical Correlation Results

Canonical R = 0.788989 Canonical R^2 = 0.622503

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.37749702	17.3148	2	21	0.0001

Raw Canonical Coefficients, Function I:
(=======⇒: Setting the larger element to one)

X1	0.1740767179	=======⇒	1
X2	0.1503001476	=======⇒	0.863413

2. Discriminant Analysis Results

Canonical R = 0.788989 Canonical R^2 = 0.622503

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.37749702	17.3148	2	21	0.0001

Raw Coefficients for Function I:
(=======⇒: Setting the larger element to one)

X1	0.2770966981	=======⇒	1
X2	0.2392489651	=======⇒	0.863413

Pooled Within-Group Covariance Matrix (S_{pooled}) DF=22

X1	5.132575758	2.412878788
X2	2.412878788	4.996212121

Between-Group Covariance Matrix (B_0) DF = 1

X1	12.08680556	11.26736111
X2	11.26736111	10.50347222

Table 5: Data for Canonical Correlation Approach to Chi-Square

ID	A	B	A1	A2	B1	B2
1	1	1	1	0	1	0
2	1	1	1	0	1	0
:	:	:	:	:	:	:
31	1	2	1	0	0	1
:	:	:	:	:	:	:
46	1	3	1	0	0	0
:	:	:	:	:	:	:
61	2	1	0	1	1	0
:	:	:	:	:	:	:
101	2	2	0	1	0	1
:	:	:	:	:	:	:
151	2	3	0	1	0	0
:	:	:	:	:	:	:
161	3	1	0	0	1	0
:	:	:	:	:	:	:
171	3	2	0	0	0	1
:	:	:	:	:	:	:
196	3	3	0	0	0	0
:	:	:	:	:	:	:
200	3	3	0	0	0	0

Data adapted from Ott, L. (1988). An Introduction to Statistical Methods and Data Analysis (3rd Ed.), Table 6.6, p.250.

Table 6: Results from Canonical Correlation Approach to Chi-Square Test

1. Chi-Square Test

Table of A by B

A Frequency (Expected)	B 1	2	3	Total
1	30 (24)	15 (27)	15 (9)	60
2	40 (40)	50 (45)	10 (15)	100
3	10 (16)	25 (18)	5 (6)	40
Total	80	90	30	200

Statistics for Table of A by B

Statistic	DF	Value	Prob
Chi-Square	4	18.194	0.001
Likelihood Ratio Chi-Square	4	18.708	0.001

Sample Size N = 200

2. Canonical Correlation Approach to Contingency Table Analysis

Canonical Correlation Analysis
Multivariate Statistics and F Approximations

Statistic	Value	F	Num DF	Den DF	Pr>F
Wilks' Lambda	0.90972222	4.7476	4	392	0.0009
Pillai's Trace	0.09097222	4.6939	4	394	0.0010

$$\chi^2 = N(\text{Pillai's Trace}) = 200 \times 0.09097222 = 18.194$$